[This Page Intentionally Left Blank]

1.0 INTRODUCTION AND RATIONALE FOR THE USE OF *IN VITRO* NEUTRAL RED UPTAKE CYTOTOXICITY TEST METHODS TO PREDICT STARTING DOSES FOR *IN VIVO* ACUTE ORAL SYSTEMIC TOXICITY TESTING

Poisoning is a more serious public health problem than is generally recognized. The Institute of Medicine estimates that more than 4 million poisoning episodes occur annually in the United States (Institute of Medicine 2004). In 2001, 30,800 deaths placed poisoning as the second leading cause of injury-related death behind automobile accidents (42,433 deaths) (Institute of Medicine 2004). The hazard potential for poisoning in humans is assessed by acute oral toxicity testing in rodents, which is a regulatory requirement for many substances and products. However, ethical and societal demands call for decreasing the numbers of animals used for such studies.

In vitro cytotoxicity methods have been evaluated as a means to reduce and refine[1] the use of animals in toxicity testing. In 1983, an international effort called the Multicentre Evaluation of *In Vitro* Cytotoxicity (MEIC) was initiated to evaluate the relationship of *in vitro* cytotoxicity to acute *in vivo* toxicity. Tests of 50 substances in 61 *in vitro* assays identified a battery of three human cell line assays that were correlated to human lethal blood concentrations. The Registry of Cytotoxicity (RC), a database that currently consists of *in vivo* acute toxicity data from rats and mice and *in vitro* cytotoxicity data from multiple cell lines for 347 substances, was published in 1998 (Halle 1998). A regression model constructed from these data was proposed by ZEBET, the German National Center for the Documentation and Evaluation of Alternative Methods to Animal Experiments, as a method to reduce animal use by identifying the most appropriate starting doses for acute oral systemic toxicity tests (Halle 1998; Spielmann et al. 1999). In October, 2000, these initiatives, a European Center for the Validation of Alternative Methods (ECVAM) testing strategy (Seibert et al. 1996), and other initiatives (ICCVAM 2001a [see Section 2.4, pg. 24])

[1] A reduction alternative is a new or modified test method that reduces the number of animals required. A refinement alternative is a new or modified test method that refines procedures to lessen or eliminate pain or distress in animals or enhances animal well-being (ICCVAM 2003).

to use *in vitro* cytotoxicity test methods to reduce animal use in acute toxicity testing were evaluated by the International Workshop on *In Vitro* Methods for Assessing Acute Systemic Toxicity (hereafter referred to as "Workshop 2000"; ICCVAM 2001a). This workshop was organized by the Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM) and The National Toxicology Program (NTP) Interagency Center for the Evaluation of Alternative Toxicological Methods (NICEATM).

ICCVAM recommended (ICCVAM 2001a) further evaluation of the use of *in vitro* cytotoxicity data as one of the factors used to estimate starting doses for *in vivo* acute lethality studies based on preliminary information that this approach could reduce the number of animals used in *in vivo* studies (i.e., reduction), minimize the number of animals that receive lethal doses (i.e., refinement), and avoid underestimating hazard. ICCVAM concurred with the Workshop recommendation that near-term validation studies should focus on two standard basal cytotoxicity assays: one using a human cell system and one using a rodent cell system. Since the murine BALB/c 3T3 cytotoxicity assay had been evaluated for only a limited number of chemical classes, there is merit in determining its usefulness with a broader array of chemical classes. A background of historical data for *in vitro* cytotoxicity testing using 3T3 cells is available through other publications (e.g., Balls et al. 1995; Brantom et al. 1997; Gettings et al. 1991, 1994a, 1994b; Spielmann et al. 1991, 1993, 1996). Human cell lines should also be considered since one of the aims of toxicity testing is to make predictions of potential toxicity in humans (ICCVAM 2001a – *ICCVAM Recommendations*). Historical data for *in vitro* cytotoxicity testing using normal human keratinocyte (NHK) cells is also available through other publications (e.g., Gettings et al. 1996; Harbell et al. 1997; Sina et al. 1995; Willshaw et al. 1994).

NICEATM, in partnership with ECVAM, designed a multi-laboratory validation study to evaluate animal reduction when using two mammalian cell types for *in vitro* basal cytotoxicity test methods with a neutral red uptake (NRU) cell viability endpoint to predict starting doses (i.e., estimated rat LD_{50} values where LD_{50} is median lethal dose) for acute oral systemic toxicity test methods. The objectives for the NICEATM/ECVAM validation study were to:

- further standardize and optimize two *in vitro* NRU cytotoxicity protocols using mouse fibroblast (BALB/c) 3T3 cells and normal human epidermal keratinocytes (NHK) in order to maximize intra- and inter-laboratory reproducibility
- refine the prediction model drawn from the ZEBET approach
- assess the accuracy of the two standardized *in vitro* basal cytotoxicity test methods for estimating rodent oral LD_{50} values across the five Globally Harmonized System of Classification and Labelling of Chemicals (GHS; United Nations [UN] 2005) categories of acute oral toxicity as well as unclassified toxicities and estimating human lethal serum concentrations
- estimate the reduction and refinement in animal use achievable from using *in vitro* basal cytotoxicity assays as one of the factors of the weight-of-evidence to identify starting doses for specific *in vivo* acute toxicity tests
- generate high quality *in vivo* lethality and *in vitro* cytotoxicity databases that can be used to support the investigation of other *in vitro* test methods necessary to improve the prediction of acute systemic toxicity

Section 1 of this background review document (BRD) summarizes the background information on the use of *in vitro* cytotoxicity test methods for predicting starting doses for acute systemic toxicity assays. It includes an overview of the correlation between *in vitro* cytotoxicity and acute lethality, the regulatory requirements for acute systemic toxicity testing, the purpose of using *in vitro* NRU assays to predict starting doses for *in vivo* acute oral systemic toxicity assays, the scientific basis of the approach, and the intended uses and applicability of this approach. **Section 2** describes the protocols used to evaluate the NRU assays using 3T3 and NHK cells. **Section 3** describes the selection of the reference substances tested in the current validation study. **Section 4** describes the derivation of reference *in vivo* rat and mouse LD_{50} values for the substances used to assess the performance of the *in vitro* NRU cytotoxicity test methods (hereafter referred to as "[3T3 and/or NHK] NRU test methods"). **Section 5** provides the 3T3 and NHK NRU data obtained during the validation study. **Section 6** refines the ZEBET approach and provides an assessment of the accuracy of the NHK and 3T3 assays for predicting acute systemic

toxicity. **Section 7** describes the assessment of the reproducibility of the assays. **Section 8** summarizes the quality of the 3T3 and NHK NRU data. **Section 9** summarizes relevant data from other studies using *in vitro* cytotoxicity test methods**. Section 10** discusses computer simulation modeling methods and results from the use of the 3T3 and NHK NRU test methods to reduce and refine animal use in acute systemic toxicity assays. **Section 11** discusses resource needs (e.g., equipment, training, time, cost) to implement these *in vitro* test methods. **Section 12** provides the references and **Section 13** provides a glossary of terms used in this BRD. The appendices provide supporting information for the aforementioned sections.

1.1 Background and Rationale for the Use of *In Vitro* Cytotoxicity Assays to Predict Starting Doses for *In Vivo* Acute Oral Systemic Toxicity Tests

Workshop 2000 was jointly sponsored by the U.S. National Institute of Environmental Health Sciences (NIEHS), the NTP, and the U.S. Environmental Protection Agency (EPA). During this workshop, participants reviewed the status of several major international *in vitro* initiatives directed toward using *in vitro* test methods to reduce the use of laboratory animals for acute toxicity testing (ICCVAM 2001a). **Sections 1.1.1** to **1.1.3** review three major initiatives evaluated by Workshop 2000 participants. **Section 1.1.4** provides information on the development of the NICEATM/ECVAM *In Vitro* NRU Cytotoxicity Validation Study.

1.1.1 The MEIC Program

The Scandinavian Society for Cell Toxicology established the MEIC program in 1983 to investigate the relevance of *in vitro* test results for predicting the acute toxicity of substances in humans (Bondesson et al. 1989). The program was an open study that invited interested laboratories worldwide to participate in testing 50 reference substances in their particular *in vitro* cytotoxicity assays. Although participating laboratories were requested to buy high purity chemicals, no effort was made to assure that all laboratories tested substances of the same purity or even purchased them from the same supplier (Clemedson et al. 1996a). Minimal methodological directives were provided to maximize protocol diversity among the 96 participating laboratories.

The reference substances were selected to represent different classes of chemicals with good data on acute toxicity (i.e., lethal doses, kinetics, and blood/serum concentrations [LC] in humans and the oral dose producing lethality in 50% of the animals [oral LD_{50} values] in rats and mice) to serve as reference values for the *in vitro* tests (Bondesson et al. 1989). The MEIC management team collected human data from clinical and forensic toxicology handbooks and case reports from human poisonings (Ekwall et al. 1998a). The data were presented and analyzed in a series of 50 MEIC Monographs. Rat and mouse oral LD_{50} data were collected from the Registry of Toxic Effects for Chemical Substances (RTECS®) from the U.S. National Institute for Occupational Safety and Health ([NIOSH]; now licensed to MDL Information Systems, Inc.).

The 50 reference substances were tested in 61 different *in vitro* assays (Ekwall et al. 1998b). The measurement of interest was the concentration producing 50% inhibition of the endpoint measured (i.e., IC_{50}, the concentration that produces 50% inhibition of the endpoint measured). Of the 20 assays that used human-derived cells, 18 used cell lines and two used primary cell cultures. Twenty-one assays used cells of animal origin (12 cell lines and nine primary cell cultures). Eighteen assays were ecotoxicological tests and two were cell-free test systems. The majority of the assays measured cell viability and/or cell growth.

The predictability of *in vivo* acute toxicity from the *in vitro* IC_{50} data was assessed against human LC values compiled from three different data sets: clinically measured acute lethal serum concentrations, acute lethal blood concentrations measured post-mortem, and peak LC values derived from approximate LC_{50} curves over time after exposure (Ekwall et al. 2000). A partial least squares (PLS) analysis indicated that the 61 assays predicted the three sets of lethal blood concentrations well (R^2 = 0.77, 0.76 and 0.83, Q^2 = 0.74, 0.72, and 0.81, respectively, where R^2 is the determination coefficient and Q^2 is the predicted variance according to cross-validation in the PLS model used). The prediction of human lethal doses by rat and mouse oral LD_{50} values with a two component PLS model was less accurate (R^2 = 0.65, Q^2 = 0.64) than the *in vitro* predictions of lethal blood concentrations.

The exposure duration for the *in vitro* assays was most often 24 hours, but ranged from 5 minutes to 6 weeks (Clemedson et al. 1996). Results suggested that basal (general) cytotoxicity can be assessed using a variety of mammalian cell lines and almost any growth/viability endpoint.

The MEIC analysis showed that the most predictive *in vitro* assays generally used human cell lines (Ekwall et al. 1998b). The MEIC study yielded a battery of *in vitro* assays with good performance for predicting acute lethality in humans (Ekwall et al. 2000). The MEIC team concluded that improvements were necessary for *in vitro* tests to be used as complete replacements for acute animal tests. To adjust for toxicity produced by mechanisms other than basal cytotoxicity, the evaluation-guided development of new *in vitro* tests (EDIT) was proposed to address targeted development of *in vitro* methods for other endpoints including biokinetics (gut absorption, distribution, clearance), biotransformation, and target organ toxicity (Clemedson et al. 2002).

1.1.2 The RC

The RC is a database of acute oral LD_{50} values for rats and mice obtained from RTECS® and IC_{50} values from *in vitro* cytotoxicity assays using multiple cell lines and cytotoxicity endpoints for substances with known molecular weights (Halle 1998). The main purpose for compiling the RC was to evaluate, with a large amount of data from substances with a wide range of systemic oral toxicities, whether basal cytotoxicity (averaged over various cells, cell lines, and/or toxicity endpoints) is a sufficiently accurate predictor of acute systemic toxicity. The RC currently contains data for 347 substances (Halle 1998) and efforts are underway to increase the number of substances to 500 (ICCVAM 2001a). To date, mixtures of chemicals have not been evaluated.

The RC includes published data for substances that met the following criteria for cytotoxicity data (Halle 1998):

- at least two different IC_{50} values were available, either from different cell types, different cell lines, or different cytotoxicity endpoints
- mammalian cells, with the exception of hepatocytes were used

- substance exposure duration was at least 16 hours, with no upper limit

The following cytotoxicity endpoints were accepted:

- cell proliferation: cell number, cell protein, DNA content, DNA synthesis, ^{3}H-thymidine intake, colony formation
- cell viability and metabolic indicators: metabolic inhibition test (MIT-24), mitochondrial reduction of tetrazolium salts into an insoluble (3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyl tetrazolium bromide [MTT]) or soluble (2,3-bis(2-methoxy-4-nitro-5- sulfophenyl)-2H-tetrazolium-5-carboxanilide [XTT]) dye
- cell viability/membrane indicators: NRU, Trypan blue exclusion, cell attachment, cell detachment
- differentiation indicators, such as functional and morphological indicators within cell clusters, and/or intracellular morphology

IC_{50} values for 347 substances were obtained from 157 original publications (Halle 1998). The 1,912 IC_{50} values, two to 32 per substance, were averaged using geometric means to produce one IC_{50x} value for each substance.

For the RC *in vivo* data, LD_{50} values published in RTECS® were used. For the first 117 substances, designated as the training data set (RC-I), LD_{50} values were not revised when subsequent issues of RTECS® reported different LD_{50} values. For the most recent 230 substances, designated as the verification set (RC-II), the LD_{50} values were taken only from the 1983/84 RTECS® publication. Whenever obtainable, oral LD_{50} data from rats were used (282 values). If rat data were unavailable, LD_{50} data from mice were used (65 values). Combining rat and mouse data in the regression was deemed to be justified when separate regressions for the mouse and rat LD_{50} data against the IC_{50x}[2] data did not result in significant differences between the slopes and intercepts of the rat and mouse regressions (Halle 1998).

[2] IC_{50x} is the geometric mean of multiple IC_{50} values collected for each substance in the RC database.

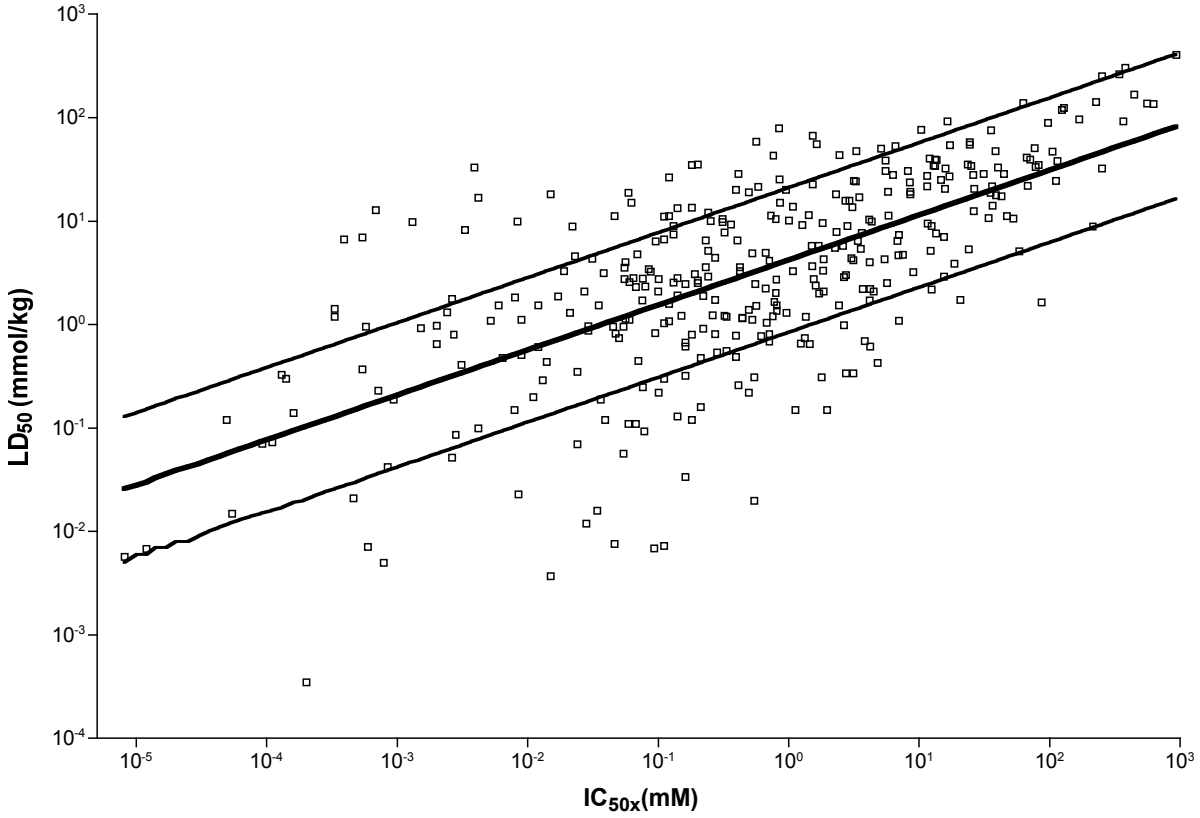
To obtain a model for the prediction of LD_{50} values from IC_{50} values, Halle (1998) calculated a linear regression from pairs of the log-transformed IC_{50x} values (in mM) and log transformed rodent oral LD_{50} values (in mmol/kg) (see **Figure 1-1**). The regression, referred to here as the "RC millimole regression," has the following formula:

$$\textbf{log } \mathbf{LD_{50}} \textbf{ (mmol/kg) = 0.435 x log } \mathbf{IC_{50x}} \textbf{ (mM) + 0.625}$$

Presumably, the substance units were expressed in moles because moles are the units that produce biological activity and, hence, are expected to produce the best fitting regression. All of the substance data were obtained for single chemicals; chemical mixtures were not included in the database and therefore were not available for determining the regression formula.

To identify an acceptability range for practical use and research purposes, the acceptable prediction interval for the LD_{50} was empirically defined as approximately one-half order of magnitude on either side of the best-fit linear regression (i.e., ± log 5, or ± 0.699) (Halle 1998). This interval was based on eight linear regressions calculated for *in vitro* cytotoxicity data, using various endpoints and mammalian cells, and *in vivo* rat, mouse, or rat and mouse LD_{50} data from five publications. It approximates the predicted LD_{50} range for the eight regressions across about eight orders of magnitude of IC_{50} values. Seventy-four percent of the RC substances fall within the prediction interval.

Figure 1-1 RC Millimole Regression Between *In Vitro* Cytotoxicity (IC_{50x}) and Rat and Mouse Acute Oral LD_{50} Values for 347 Chemicals



The heavy line shows the fit of the data to a linear regression model, log (LD_{50}) = 0.435 x log (IC_{50x}) + 0.625; r=0.67. IC_{50x} values are the geometric means of multiple endpoints and cell types. The thinner lines show the empirical prediction interval (± log 5, or ± 0.699) that is based on the anticipated precision for the prediction of LD_{50} values from cytotoxicity data (Halle 1998).

1.1.3 The ZEBET Initiative to Reduce Animal Use

The concept that the predicted LD_{50} value could be used as a starting dose for acute oral toxicity testing to reduce the number of animals was first discussed at an ECVAM workshop (Seibert et al. 1996) as it related to the, then new, sequential dosing methods such as the Acute Toxic Class method (ATC; OECD draft TG 423 [ICCVAM 2001a]) and the Up-and-Down Procedure (UDP; OECD draft TG 425 [ICCVAM 2001a]). In these tests, for which the OECD guidelines have now been finalized, the number of animals needed depends upon the choice of the starting dose, since the number of consecutive dosing steps (and thus the number of animals used) is reduced as the starting dose more closely approximates the true

toxicity class (ATC or Fixed Dose Procedure [FDP]), or the true LD_{50} (UDP). The ZEBET approach involves using an IC_{50} value from an *in vitro* basal cytotoxicity test to predict an LD_{50} close to the true LD_{50}. The IC_{50} is used in the RC millimole regression to predict an LD_{50} value for use as a starting dose for the ATC or UDP (Spielmann et al. 1999). The use of *in vitro* cytotoxicity assays to predict a starting dose equivalent to the LD_{50} may reduce animal use in the UDP by 25-40%, depending upon the slope of the curve and the stopping rule applied (Spielmann et al. 1999; ICCVAM 2001a).

1.1.4 The NICEATM/ECVAM *In Vitro* NRU Cytotoxicity Validation Study

Workshop 2000 participants concluded that none of the *in vitro* models reviewed had been formally evaluated for reliability and relevance, and their usefulness and limitations for generating information for acute toxicity testing had not been assessed. However, the approach proposed by ZEBET (Halle 1998; Spielmann et al. 1999) was recommended for rapid adoption so that data could be generated to establish its usefulness with a large number of substances (ICCVAM 2001a). To assist in the adoption and implementation of the ZEBET approach, several workshop participants wrote *Guidance Document on Using In Vitro Data to Estimate In Vivo Starting Doses for Acute Toxicity* (hereafter referred to as *Guidance Document*; ICCVAM 2001b).

The *Guidance Document* recommended testing 10 to 20 reference substances of high purity from the RC in a candidate *in vitro* basal cytotoxicity assay to be used for predicting starting doses for acute oral lethality tests (ICCVAM 2001b). The substances were to cover a wide range of toxicity and fit the RC prediction model (i.e., the linear regression line) as closely as possible. The assays recommended and provided as examples are NRU assays using 3T3 and NHK cells. The IC_{50} results for the selected substances would be used to calculate a new regression line with the LD_{50} values used by the RC. If the resulting regression were parallel to the RC millimole regression and within the ± log 5 (i.e., ± 0.699) prediction interval for the RC, the *Guidance Document* recommended using the cytotoxicity assay to predict starting doses for LD_{50} assays. If the regression from the assay did not meet these criteria, then the *Guidance Document* advised either (a) adjusting the slope or (b) using the NRU protocols offered in the *Guidance Document* (considered the most efficient approach).

To further characterize the usefulness of the 3T3 and NHK NRU test methods as predictors of starting doses for acute oral systemic toxicity assays, NICEATM and ECVAM designed an independent[3] multi-laboratory validation study to evaluate the performance of these *in vitro* test methods. The inclusion of human cells in the NICEATM/ECVAM validation study implements a Workshop 2000 recommendation to evaluate whether cytotoxicity in human or rodent cells best predicts human acute toxicity. ECVAM's development of a prediction model for human acute toxicity using data collected in the NICEATM/ECVAM validation study will be addressed elsewhere.

Study Design

The planning phases of the NICEATM/ECVAM validation study included the selection of reference substances for testing, which is described in **Section 3,** and the identification of reference LD_{50} values for the reference substances, which is described in **Section 4**. The NRU testing proceeded in several phases (See **Figure 1-2**) so that the Study Management Team (SMT) could evaluate the reproducibility of results after each phase and refine the protocols, if necessary, before proceeding to the next phase. The NRU data collected during the laboratory phase were used to evaluate, and in some cases, develop, linear regression formulas for the prediction of LD_{50} values by IC_{50} values (see **Section 6**). Computer simulation modeling of acute oral toxicity test outcomes was then performed to determine animal savings using the NRU-predicted starting doses compared with the default starting dose (see **Section 10**). Study management and study participant information is provided in **Appendix A.**

[3] "Independent" is used here to indicate that neither NICEATM nor ECVAM neither developed nor had monetary interest in the test methods.

Figure 1-2 NICEATM/ECVAM Validation Study Phases

Phase Ia: Laboratory Evaluation

Development of a positive control database for each laboratory

- Perform at least 10 replicate NRU tests of the positive control substance (sodium laurel sulfate [SLS]) with each cell type.
- Calculate mean $IC_{50} \pm 2$ SD for each cell type for each lab.
- Establish acceptance criteria for positive control performance in future assays.

⇓

Phase Ib: Laboratory Evaluation

Limited substance testing to demonstrate the reliability of the protocol

- Each laboratory tests the same three coded substances of varying toxicities three times with each cell type.
- Refine protocols and repeat, if necessary, until acceptable intra/interlaboratory reproducibility is achieved.

⇓

Phase II: Laboratory Qualification

Evaluation of protocol refinements

- Each laboratory tests nine coded substances covering the range of GHS toxicity categories, with three replicate tests/substance for each test method.
- Assure that corrective actions taken in Phase I have achieved the desired results.
- Further refine protocols and re-test, if necessary, to achieve acceptable reliability.
- Finalize protocols for Phase III.

⇓

Phase III: Laboratory Testing Phase

Test of optimized protocols

- Each laboratory tests 60 coded substances three times using the final protocol for each test method.

1.2 Regulatory Rationale and Applicability for the Use of *In Vitro* Cytotoxicity Test Methods to Predict Starting Doses for Acute Oral Systemic Toxicity Testing

1.2.1 Current Regulatory Testing Requirements for Acute Systemic Toxicity

The major regulatory requirement for acute systemic toxicity testing is for the hazard classification and labeling of products, which is intended to protect handlers and consumers from toxic hazards. The LD_{50} results from acute systemic toxicity tests are used to place substances in various toxicity categories that, in turn, invoke the associated hazard phrases to be used on product labels. **Table 1-1** shows the current U.S. legislation requiring the use of acute systemic toxicity testing for product labeling and the substances regulated. **Table 1-2** shows the statutory protocol requirements and classification systems used by each U.S. regulatory agency. Also included is an international guideline for labeling, the Harmonized Integrated Classification System for Human Health and Environmental Hazards of Chemical Substances and Mixtures, which provides guidance to regulatory agencies on the use of the GHS (UN 2005) as a method for an internationally comprehensible system for hazard communication (OECD 2001b).

Table 1-1 Summary of Current U.S. Legislation for Using Acute Systemic Toxicity Data for Product Labeling

Legislation (Year of Initial Enactment)	U.S. Regulatory Agency	Substance
Federal Insecticide, Fungicide and Rodenticide Act (1947)	EPA	Pesticides
Federal Hazardous Substances Act (1964)	CPSC	Household products
Occupational Safety and Health Act (1970)	OSHA	Occupational materials
Federal Hazardous Material Transportation Act (1975)	DOT	Transported substances

Abbreviations: EPA = U.S. Environmental Protection Agency; CPSC = U.S. Consumer Product Safety Commission; OSHA = U.S. Occupational Safety and Health Administration; DOT = U.S. Department of Transportation.

Note: The U.S. Food and Drug Administration (FDA) does not require data for acute lethality testing, and in fact, discourages the use of animals for such testing (FDA 1993).

412 **Table 1-2 Regulatory Classification Systems for Acute Oral Toxicity**

413

Regulatory Agency (Authorizing Act)	Animals	Endpoint	Classification
EPA (Federal Insecticide, Fungicide and Rodenticide Act)	Use current EPA or OECD protocol	Death[1]	I - $LD_{50} \leq 50$ mg/kg II - $50 < LD_{50} \leq 500$ mg/kg III – $500 < LD_{50} \leq 5000$ mg/kg IV - $LD_{50} > 5000$ mg/kg
CPSC (Federal Hazardous Substances Act)	White rats, 200-300 g	Death[1] within 14 days for ≥ half of a group of ≥ 10 animals	Highly toxic – $LD_{50} \leq 50$ mg/kg Toxic – 50 mg/kg $< LD_{50} <$ 5 g/kg
OSHA (Occupational Safety and Health Act)	Albino rats, 200-300 g	Death[1], duration not specified.	Highly toxic - $LD_{50} \leq 50$ mg/kg Toxic – $50 < LD_{50} < 500$ mg/kg
DOT (Federal Hazardous Material Transportation Act)	Male and female young adult albino rats	Death[1] within 14 days of half the animals tested. Number of animals tested must be sufficient for statistically valid results.	Packing Group 1 - $LD_{50} \leq 5$ mg/kg Packing Group II – $5 < LD_{50} \leq 50$ mg/kg Packing Group III – $LD_{50} < 500$ mg/kg (liquid) $LD_{50} < 200$ mg/kg (solid)
OECD Guidance for Use of GHS (2001a)	Protocol not specified	Protocol not specified	I - $LD_{50} \leq 5$ mg/kg II - $5 < LD_{50} \leq 50$ mg/kg III - $50 < LD_{50} \leq 300$ mg/kg IV - $300 < LD_{50} \leq 2000$ mg/kg V - $2000 < LD_{50} \leq 5000$ mg/kg Unclassified - $LD_{50} > 5000$ mg/kg

414 [1]*Guidance Document on the Recognition, Assessment and Use of Clinical Signs as Humane Endpoints*
415 *for Experimental Animals Used in Safety Evaluation* calls for humane killing of moribund animals
416 (OECD 2000). Moribund animals that are humanely euthanized are accepted as deaths.
417 Abbreviations: EPA = U.S. Environmental Protection Agency; CPSC = U.S. Consumer Product Safety
418 Commission; OECD = Organisation for Economic Co-operation and Development; OSHA = U.S.
419 Occupational Safety and Health Administration; DOT = U.S. Department of Transportation; GHS =
420 Globally Harmonized System of Classification and Labelling of Chemicals (UN 2005)

421

In addition to classification and labeling, acute systemic toxicity test results may be used for:

- establishing dosing levels for repeated dose toxicity studies
- generating information on the specific organs affected
- providing information related to the mode of toxic action
- aiding in the diagnosis and treatment of toxic reactions
- providing information for comparison of toxicity and dose response among substances in a specific chemical or product class
- aiding in the standardization of biological products
- aiding in judging the consequences of single, high accidental exposures in the workplace, home, or from accidental release
- serving as a standard for evaluating alternatives to animal tests

Test Methods for Assessing Acute Systemic Toxicity

The current internationally recognized test methods for acute systemic toxicity testing are the FDP (OECD 2001c), the ATC method (OECD 2001d), and UDP (OECD 2001a; EPA 2002a) (see **Appendix M** for test method guidelines). Information on signs of acute toxicity and target organs can be obtained using any of the three test methods. All three methods are sequential tests in which the outcome of testing one or more animals at the first dose is used to determine the second dose that should be tested. The FDP differs from the UDP and ATC in that it involves testing more animals per dose and the primary endpoint of interest is evident toxicity[4] rather than lethality. Both the FDP and the ATC method provide a range for the LD_{50} for classification purposes. The UDP generally provides a point estimate of the LD_{50} with a confidence interval (EPA 2002a).

Each of the test method guidelines include a limit test in which up to five (UPD and FDP) or six (ATC) animals are tested at the limit, or upper bound, dose (OECD 2001a,c,d; EPA 2002a). The limit test can be performed using 2000 or 5000 mg/kg.

[4] *Evident toxicity* is a general term describing clear signs of toxicity following administration of test substance, such that an increase to the next highest fixed dose would result in the development of severe toxic signs and probably mortality (ICCVAM 2000).

1.2.2 Intended Regulatory Uses for the *In Vitro* Cytotoxicity Test Methods

In vitro cytotoxicity test methods are not recommended for the replacement of acute oral toxicity tests in animals. Rather, such test methods are intended to serve as adjuncts for *in vivo* acute systemic toxicity test methods. To select a starting dose for a test substance, the current test guidelines for acute oral systemic toxicity recommend using information on structurally-related substances and the results of any other toxicity tests (EPA 2002b), including *in vitro* cytotoxicity results (OECD 2001a, c, d; EPA 2002a). The 3T3 and NHK NRU test methods are intended to be used as part of the weight-of-evidence approach to select starting doses for the UDP and ATC assays in order to reduce and refine the use of animals for *in vivo* acute toxicity testing. The reduction of animals achievable with the use of basal cytotoxicity as an adjunct to the UDP or ATC is provided in **Section 10**. Since the estimation of the true LD_{50} is irrelevant to setting doses for measuring evident toxicity, the FDP will not be considered further in this document.

Section 10 presents analyses that characterize the extent of animal reduction and refinement that may occur by using the NRU test methods to estimate the starting doses for the UDP and the ATC method. Animal use and animal deaths for UDP and ATC testing is determined using computer simulation techniques rather than by animal testing. The simulations of UDP and ATC testing determine the number of animals used when using the default starting dose and when using a starting dose determined from the NRU test methods. The number of animals used with the NRU-determined starting dose is compared with the number of animals used with the default starting dose to determine the reduction in animal use with the NRU-determined starting dose. To characterize the extent of refinement produced by using the NRU-determined starting dose, the number of animals that die with the NRU-determined starting dose is compared with the number of animals that die when using the default starting dose.

1.2.3 Similarities and Differences in the Endpoints of the *In Vitro* Cytotoxicity Test Methods and *In Vivo* Acute Oral Toxicity Test Methods

The endpoint measured in the *in vitro* NRU cytotoxicity test methods is cell death (neutral red [NR] is taken up only by live cells) and the major endpoint of interest is the concentration

at 50% inhibition of NRU (i.e., the IC_{50}). The endpoint measured in acute systemic toxicity assays is usually animal death. Cell death and animal death may be similar since animals are comprised of organ systems consisting of tissues, which are comprised of cells. All cells, regardless of whether they are in animals or *in vitro* cell cultures, have similar cellular mechanisms of energy production and utilization and maintenance of cell membrane integrity.

Animal death and death of cells in culture due to toxicity are similar in that both involve some type of cellular injury. For the animal, the cellular injury produces tissue and organ injury to the most sensitive target organ, which may then cause the death of the whole organism. Organ system failure can be due either to the death of cells in the affected organ or to the loss of function of the surviving cells in the organ, which results in cell death or loss of function in other organs (Gennari et al. 2004). Death of an animal is produced by major organ system failure. Ultimately the cardiovascular and respiratory systems fail. Respiratory depression may be due to depression of the central nervous system (CNS) rather than a direct assault on the respiratory system. Other major organ system failures, such as liver and kidney failure, gastrointestinal corrosion, and bone marrow depression, also produce death. Cell death in a culture system involves the death of a single cell type. Cell death and animal death may be produced by the same mechanisms, such as disruption of membrane structure or function, inhibition of mitochondrial function, disturbance of protein turnover, disruption of energy production, etc. (Gennari et al. 2004).

Animal and cell culture systems are different with respect to how a substance or toxin is delivered to the cell and how it is distributed, metabolized, and excreted. After oral administration, animals must absorb the toxin from the gastrointestinal tract, which involves the passage of membranes. The toxin may or may not be heavily bound to serum proteins; this would reduce the availability of the toxin to the target organ. The toxin may then be metabolized during and/or after distribution to the target organs and then the toxin or its metabolites are excreted. In a cell culture system, the only membranes that must be passed are those of the target cell and cellular organelles. No absorption and distribution by other cellular systems is required. Cell culture systems may or may not include serum proteins,

which could reduce the availability of toxin to act as its target site. The 3T3 cell culture system includes serum while the NHK cell culture system does not. The 3T3 and NHK cell culture systems have little to no capacity to metabolize xenobiotic compounds. Excretion from the cell culture milieu cannot occur since cell culture systems have no excretory system. The cultured cells are exposed to substances for the entire duration of exposure in the test system.

Animal and cell culture systems may also be different with respect to the target on which a toxin acts. If a toxin acts in a specialized organ system in a whole animal, it may not produce a toxic effect by the same mechanism in cultured cells that are derived from tissue different from the target organ. For example, a neurotoxin that acts by a neuroreceptor-mediated pathway in animals, would be expected to produce toxicity by a different mechanism in 3T3 or NHK cells, which are derived from fibroblasts, and skin cells, respectively. Even if a neurotoxin were applied to neuronal cells in culture, the cultured cells may not respond in the same way as neuronal cells in a whole animal. Cultured cells may not retain the same functionality as cells *in vivo*.

1.2.4 Use of *In Vitro* Cytotoxicity Test Methods in the Overall Strategy of Hazard Assessment

In the overall strategy of hazard or safety assessment, the intended regulatory use of *in vitro* test methods is to reduce and refine the use of animals in current acute systemic toxicity assays (i.e., serve as adjuncts to these test methods). *In vitro* cytotoxicity test methods are not intended as replacements for the *in vivo* tests. For current OECD acute systemic toxicity assays (the ATC or UDP), that use sequential dosing methods, the number of animals used depends on the choice of starting dose since the number of dosing steps (and animals) is reduced if the starting dose is close to the true toxicity class (ATC) or to the true LD_{50} (UDP) (Spielmann et al. 1999; ICCVAM 2001b).

As noted earlier, Spielmann et al. (1999) and the *Guidance Document* (ICVAM 2001b) suggest that the RC millimole regression be used with *in vitro* cytotoxicity data to predict starting doses for the ATC and UDP. The approach can be applied to substances with purity

appreciably lower than 100% as long as molecular weight and purity are known. Therefore, this approach is not applicable to mixtures such as product formulations or unknown substance samples.

Thus, in addition to evaluating the reduction of animal use associated with the ATC and UDP when the current RC millimole regression (in millimolar units) is used to predict the starting dose, this study also evaluated the reduction in animal use associated with regressions based on weight units.

1.3 Scientific Basis for the *In Vitro* NRU Test Methods

Cytotoxicity has been defined as the adverse effects resulting from interference with structures and/or processes essential for cell survival, proliferation, and/or function (Ekwall 1983). Ekwall (1983) described the concept of "basal cell functions" that virtually all cells possess (mitochondria, plasma membrane integrity, etc.) and suggested that, for most substances, toxicity is a consequence of non-specific alterations in those cellular functions, which may then lead to effects on organ-specific functions and/or death of the organism. These effects may involve the integrity of membranes and the cytoskeleton, cellular metabolism, the synthesis and degradation or release of cellular constituents or products, ion regulation, and cell division.

Ekwall (1983) and others (Grisham and Smith 1984) concluded that, since the actions of substances that produce injury and death are ultimately exerted at the cellular level, *in vitro* cytotoxicity assays may be useful for the prediction of acute lethal potency. Considerable research has been undertaken to develop and evaluate *in vitro* tests for use as screens and as potential replacements for LD_{50} tests. Good agreement between cytotoxicity *in vitro* and animal lethality have been reported by numerous groups (see reviews by Phillips et al. 1990; Garle et al. 1994; Guzzie 1994). However, none of the proposed *in vitro* models have been evaluated in any formal studies for reliability and relevance, and their usefulness and limitations for generating information to meet regulatory requirements for acute toxicity testing have not been assessed.

1.3.1 Purpose and Mechanistic Basis of the *In Vitro* NRU Test Methods

There are a number of basal cytotoxicity endpoints that measure cell death and or cell proliferation. The NRU test methods were chosen for the NICEATM/ECVAM validation study because they were recommended in the *Guidance Document* for the purpose of obtaining cytotoxicity information to predict starting doses for acute systemic toxicity assays (ICCVAM 2001b). Both the 3T3 and NHK NRU test methods were reproducible in previous validation studies (ICCVAM 2001b). In addition, both cell types are easily obtainable from commercial sources and the *Guidance Document* provided preliminary evidence that these assays could reproduce the RC millimole regression. Additionally, the assays can be automated and they require no radioactivity or highly dangerous substances (see **Section 2** for the protocols).

Neutral red is a weakly cationic water-soluble dye that stains living cells (Borenfreund and Puerner 1985). It readily diffuses through the plasma membrane and concentrates in lysosomes where it electrostatically binds to the anionic lysosomal matrix. Toxins can alter the cell surface or the lysosomal membrane seeming to cause lysosomal fragility and other adverse changes that gradually become irreversible. Thus, cell death and/or inhibition of cell growth decreases the amount of neutral red taken up by the culture. The protocol for the NRU assay using 3T3 cells was first published by Borenfreund and Puerner (1985) as a two component test for toxicity screening that was standardized for a 96-well plate format. The two components were (1) a morphological examination of the cells under an inverted phase microscope and (2) a quantitative measurement of NRU. The morphological examination was designed to identify the highest tolerated dose for the assay (i.e., the highest concentration of toxicant that the cells can tolerate and that causes minimal morphological changes). This concentration was comparable to the quantitative measurement of 10% inhibition (i.e., NR_{90} value compared to the controls) of NRU. The NR_{90} value is the point where a test compound produces a significant toxic effect. The assay was said to be a rapid, reliable, inexpensive, and reproducible *in vitro* assay for screening potentially toxic agents, and it was suggested that the test was a good candidate for inclusion in a battery of tests for toxicity screening for the purpose of reducing the use of animals for toxicity tests.

1.3.2 Similarities and Differences in the Modes/Mechanisms of Action for the *In Vitro* NRU Test Methods Compared with the Species of Interest

Although the ultimate species of interest for acute systemic toxicity concerns is humans, labeling and hazard identification requirements are based on rodent studies. There are differences between humans and rodents in terms of absorption, distribution, metabolism, excretion, and the intrinsic sensitivity of target organs to xenobiotic compounds. The differences are largely substance specific. *In vitro* cytotoxicity studies have also noted differences in sensitivity between human cells and other mammalian cells (Clemedson et al. 1996).

Ekwall et al. (1998b) showed that *in vitro* cytotoxicity methods using human cell lines generally predicted human toxicity better than methods using other mammalian cell types. **Section 6** shows that, for the reference substances tested in this study, the 3T3 NRU test method usually predicted rodent acute toxicity better than the NHK NRU test method did. A human cell type, such as the NHK, may predict human toxicity better than 3T3 cells, which originate in mice (this evaluation is not reported in this BRD, but will be reported elsewhere).

Besides the species differences, there are several other differences between the 3T3 and NHK cells.

- The 3T3 cells are an immortal line, while the NHK cells are primary cells.
- They originate from different tissues; 3T3 cells are derived from embryonic fibroblasts, while the NHK cells come from neonatal foreskin tissue.
- NHK cells grow more slowly in culture than the 3T3 cells.
- NHK cells have greater ability to metabolize xenobiotic compounds, in that they exhibit some cytochrome P450 activity (Babich et al. 1991). 3T3 cells have practically no ability to metabolize xenobiotic compounds (INVITTOX 1991).

1.3.3 Range of Substances Amenable to the *In Vitro* NRU Test Methods

The *in vitro* NRU test methods can be applied to a wide range of substances as long as the substances can be dissolved in the cell culture medium or in a solvent that can be mixed with culture medium. Although these test methods may to be applicable to mixtures, none were

evaluated in this validation study. The toxicity of substances with specific mechanisms of toxicity not expected to be active in 3T3 or NHK cells (e.g., those that are neurotoxic, cardiotoxic, interfere with energy utilization, or alkylate proteins and other macromolecules) will likely be underpredicted by these test methods. Therefore, until a more predictive approach is developed, the results from basal cytotoxicity testing with such substances may not be appropriate.

Insoluble substances or those unstable or explosive in water are not compatible with the test system. Volatile substances may yield acceptable results if CO_2 permeable plastic film is used to seal the test plates. Testing for corrosive substances is unnecessary since there is no regulatory requirement for acute systemic toxicity testing for corrosives. The toxicity of substances that are highly bound to serum proteins may be underestimated by the 3T3 assay since the culture medium contains 5% serum during substance exposure. The toxicity of substances that specifically affect lysosomes may be overestimated since they may affect NRU. Red substances that absorb light in the optical density range of NR may interfere with the test if they remain inside the cell in sufficient amounts after washing and are soluble in the NR solvent.